Introduction to Machine Learning

Tushar B. Kute,
http://tusharkute.com

Machine Learning

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.
- The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.
- The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

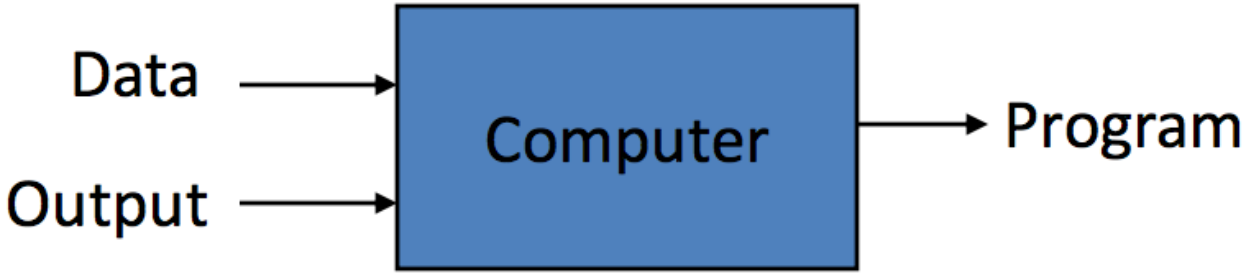
Origins of Machine Learning

- The earliest databases recorded information from the observable environment.
- Astronomers recorded patterns of planets and stars; biologists noted results from experiments crossbreeding plants and animals; and cities recorded tax payments, disease outbreaks, and populations. Each of these required a human being to first observe and second, record the observation.
- Today, such observations are increasingly automated and recorded systematically in ever-growing computerized databases.

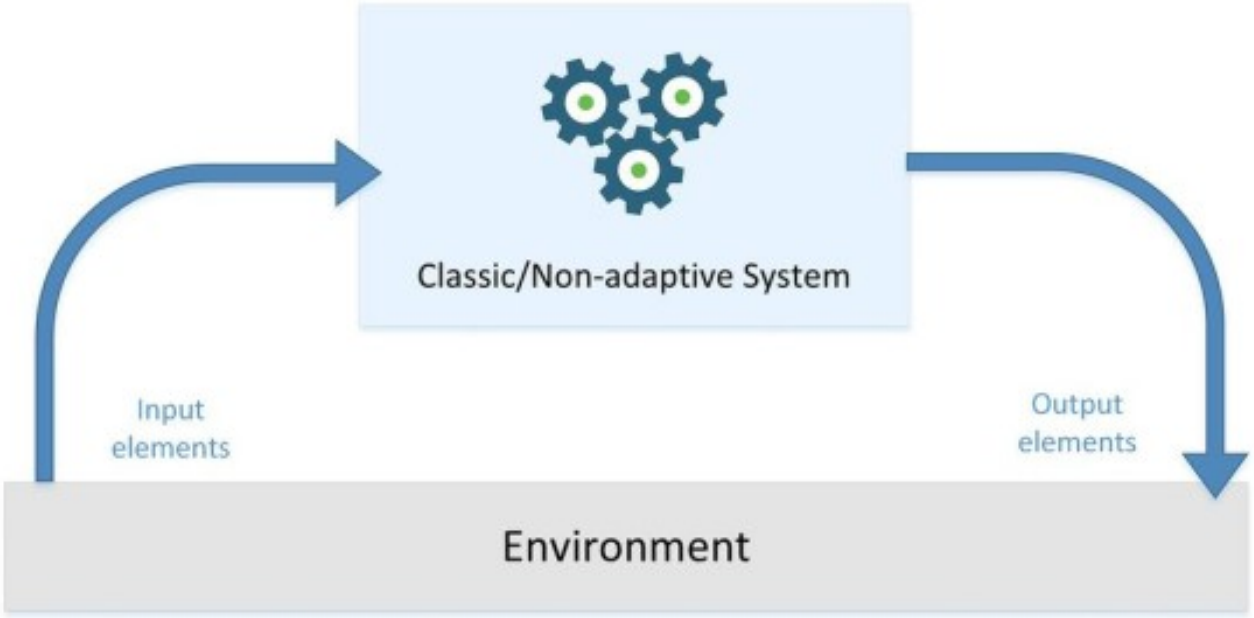
Machine Learning

Traditional Programming

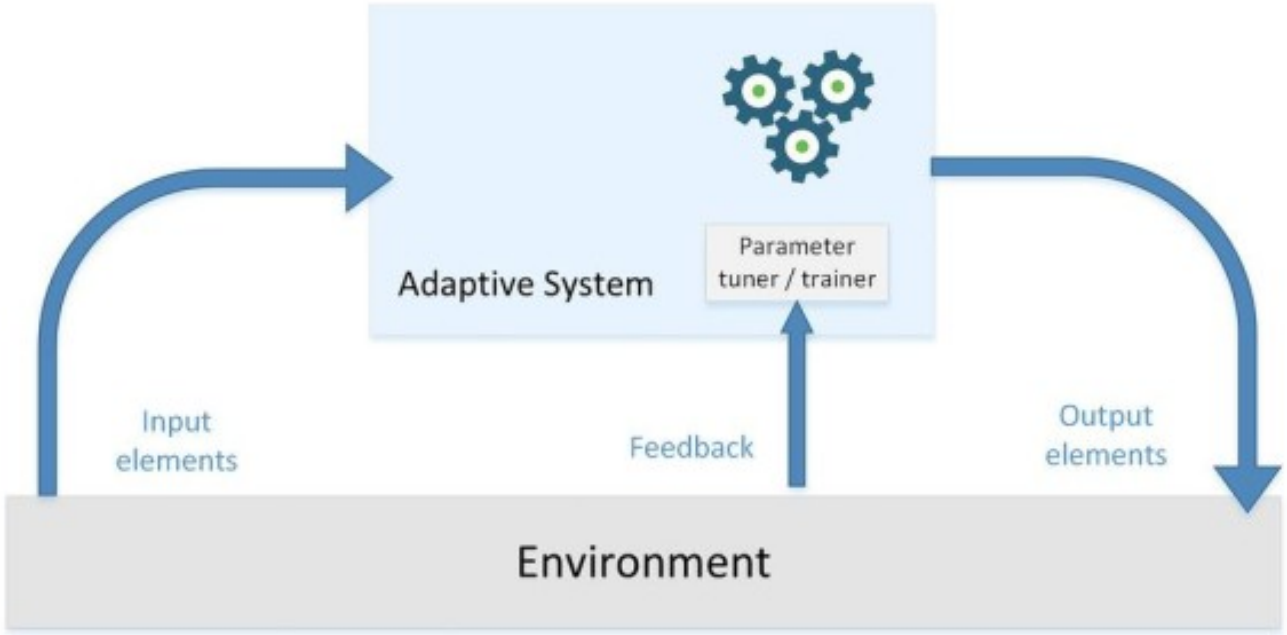
Machine Learning



Classic Systems

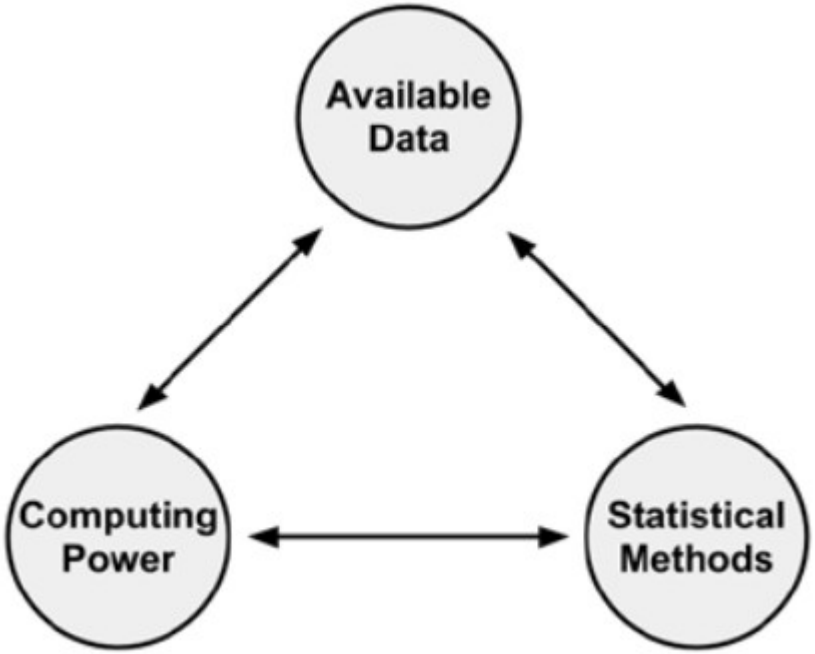


Adaptive Systems



Machine Learning

- The field of study interested in the development of computer algorithms for transforming data into intelligent action is known as machine learning.



Uses and Abuses

- Predict the outcomes of elections
- Identify and filter spam messages from e-mail
- Foresee criminal activity
- Automate traffic signals according to road conditions
- Produce financial estimates of storms and natural disasters
- Examine customer churn
- Create auto-piloting planes and auto-driving cars
- Stock market predition
- Target advertising to specific types of consumers

Case Study

सकाळ

विद्यापीठात विद्यार्थ्यांचा 'एक्झिट पोल'

'रँडम फॉरेस्ट मॉडेल'नुसार युतीच राज्यात आघाडीवर

पुणे, ता. २१ : राज्यात भाजप आणि शिवसेना युती आघाडीवर असेल, असा अंदाज वर्तविणाऱ्या चाचण्यांचे कल (एक्झिट पोल) नुकतेच प्रसिद्ध झाले आहेत. सावित्रीबाई फुले पुणे विद्यापीठातील विद्यार्थ्यांनीही त्याला दुजोरा दिला आहे. भारतीय जनता पक्षाला १७ ते २३ आणि शिवसेनेला १६ ते २१ जागा मिळतील, असा अंदाज विद्यार्थ्यांनी 'रँडम फॉरेस्ट मॉडेल' पद्धत वापरून वर्तविला आहे. राष्ट्रवादी काँग्रेसला ३ ते ९ व काँग्रेसला १ ते ६ जागा मिळतील, असा अंदाज त्यांनी वर्तवला आहे.

विद्यापीठाच्या संख्याशास्त्र विभागातील एमएस्सी (द्वितीय वर्ष)

करणारे विनय तिवारी, आर. विश्वनाथ, शरद कोळसे या विद्यार्थ्यांनी सहायक प्राध्यापक डॉ. आकांक्षा काशीकर यांच्या मार्गदर्शनाखाली हा अंदाज दिला आहे.

निवडणूक आयोगाच्या संकेतस्थळावरून सर्वेक्षणासाठी लागणारी माहिती त्यांनी मिळविली. जनमानसाचा कल ओळखण्यासाठी 'सीएसडीएस-लोकनीती' सर्वेक्षण अहवालातून नोंदी घेतल्या. त्याचबरोबर सध्याच्या सरकारच्या कामगिरीबद्दल लोकांच्या प्रतिक्रिया, पंतप्रधानपदाच्या संभाव्य उमेदवारांची लोकप्रियता, मागील निवडणुकीतील आपले मत यंदा बदलू इच्छिणारे मतदार यांचा अभ्यास करण्यात आला. या अंदाजासाठी रँडम फॉरेस्ट मॉडेल वापरण्यापूर्वी २००९ आणि २०१४च्या निवडणुकांचे अंदाज पडताळून पाहण्यात आले. हे अंदाज प्रत्यक्ष निकालांशी पडताळून पाहिले असता, ते जवळपास ९६ टक्के जुळत असल्याचे निदर्शनास आले. म्हणूनच अभ्यासात माहितीच्या विश्लेषणासाठी या पद्धतीचा वापर करण्यात आला, असे डॉ. काशीकर यांनी सांगितले.

> संख्याशास्त्र आणि संगणकशास्त्र याची सांगड घालून आणि मशिन लर्निंगच्या साह्याने उपलब्ध माहितीचे विश्लेषण केले. संख्याशास्त्रातील अभ्यासाची वेगवेगळी मॉडेल वापरून १९७७ पासून ते आतापर्यंतच्या लोकसभा आणि विधानसभा निवडणुकीतील माहितीचा अभ्यास केला. त्यामुळे संख्याशास्त्राचा वापर करून वर्तविलेला अंदाज हा निवडणुकीच्या निकालांच्या जवळ जाणारा असेल.
> - शरद कोळसे, विद्यार्थी

Pune, Main
22/05/2019 Page No. 4

Recognizing patterns

- Pattern recognition is the automated recognition of patterns and regularities in data. It has applications in
  - statistical data analysis,
  - signal processing,
  - image analysis,
  - information retrieval,
  - bioinformatics,
  - data compression,
  - computer graphics and
  - machine learning.

How do machine learn ?

- A commonly cited formal definition of machine learning, proposed by computer scientist Tom M. Mitchell, says that a machine is said to learn if it is able to take experience and utilize it such that its performance improves up on similar experiences in the future.
- This definition is fairly exact, yet says little about how machine learning techniques actually learn to transform data into actionable knowledge.

Training a dataset

- The process of fitting a particular model to a dataset is known as training.
- Why is this not called learning? First, note that the learning process does not end with the step of data abstraction.
- Learning requires an additional step to generalize the knowledge to future data.
- Second, the term training more accurately describes the actual process undertaken when the model is fitted to the data.

Practical Machine Learning

X	Y	Z
5	2	14
8	5	22
4	8	14
9	2	20
7	1	15
7	8	23

Inputs

Output

Z = ?

---> ML Model

Practical Machine Learning

X	Y	Z	Pre	Error
5	2	14	12	-2
8	5	22	21	-1
4	8	14	16	+2
9	2	20	20	0
7	1	15	15	0
7	8	23	22	-1

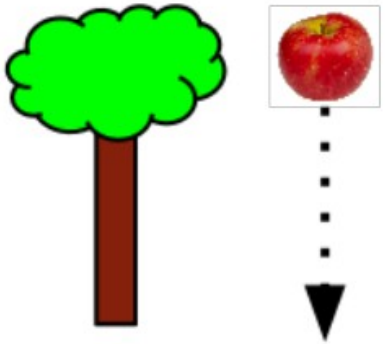
```
Z = 2X + Y                  ---> ML Model

Prediction --->   X = 6   Y = 8   Z = ?

if 20 == 19:                95%
```

Training a dataset

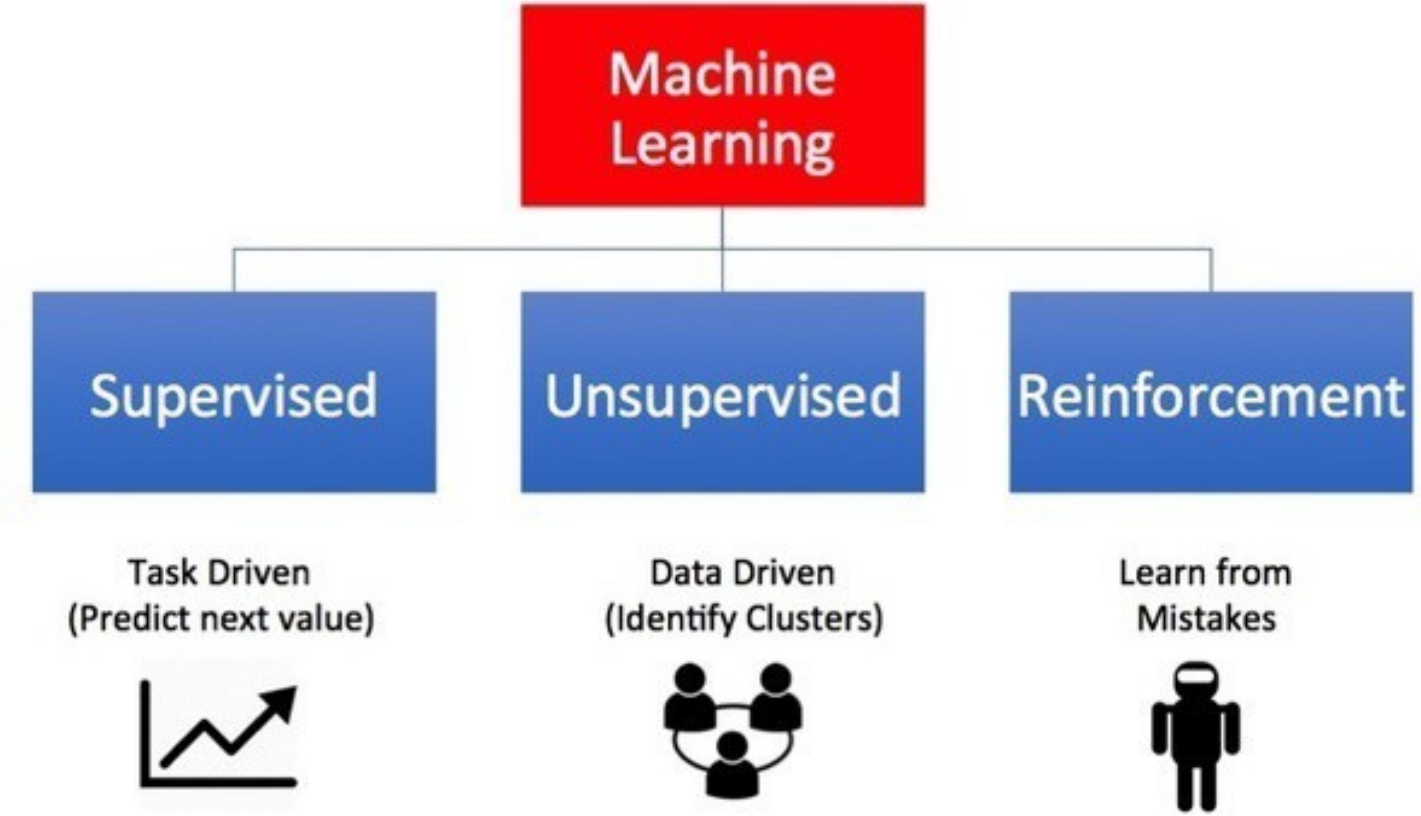
Observations → **Data** → **Model**



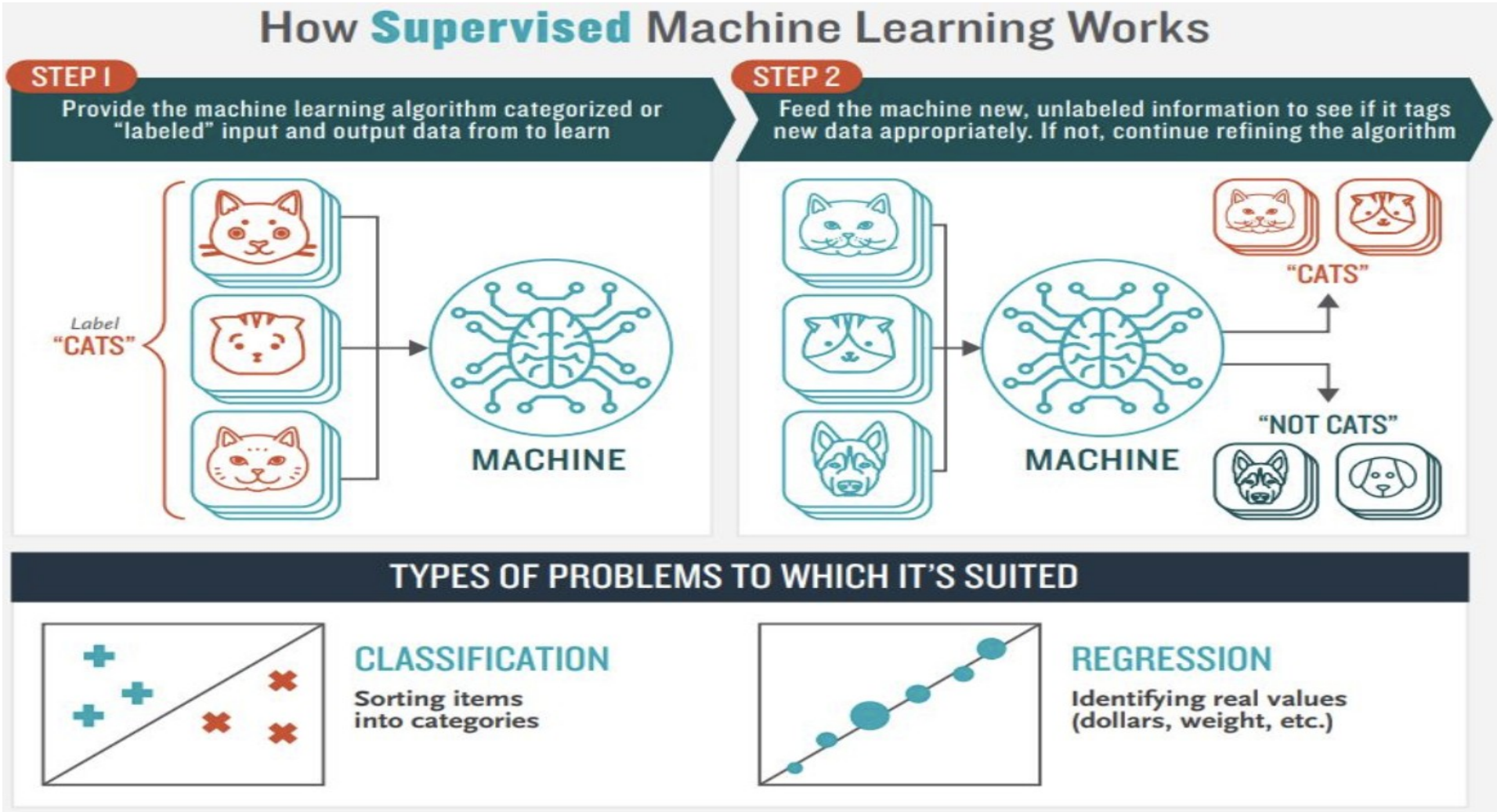
velocity	time
9.8	1
39.2	2
88.2	3
156.8	4
245	5

$$g = 9.8\ m/s^2$$

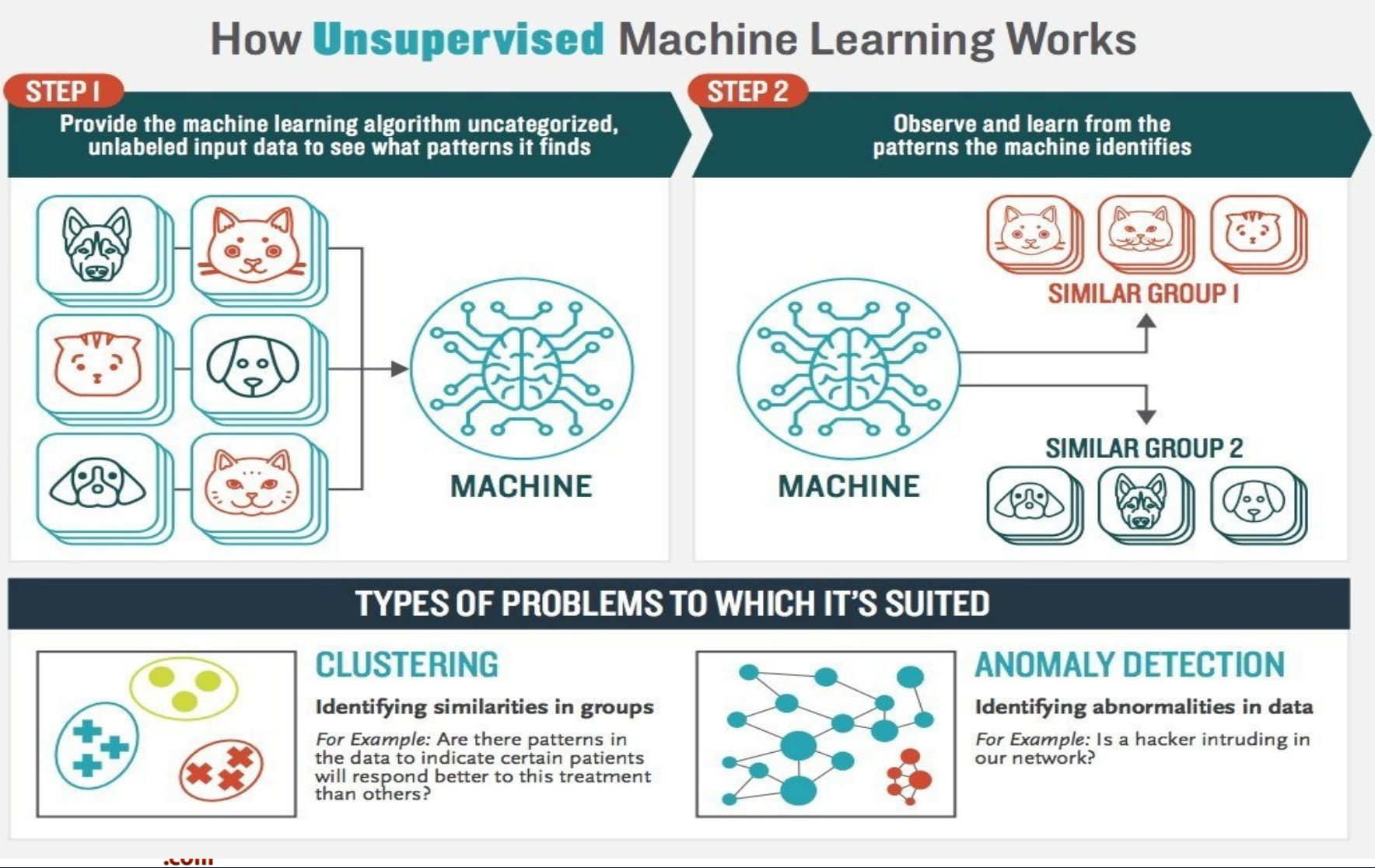
Types of Machine Learning



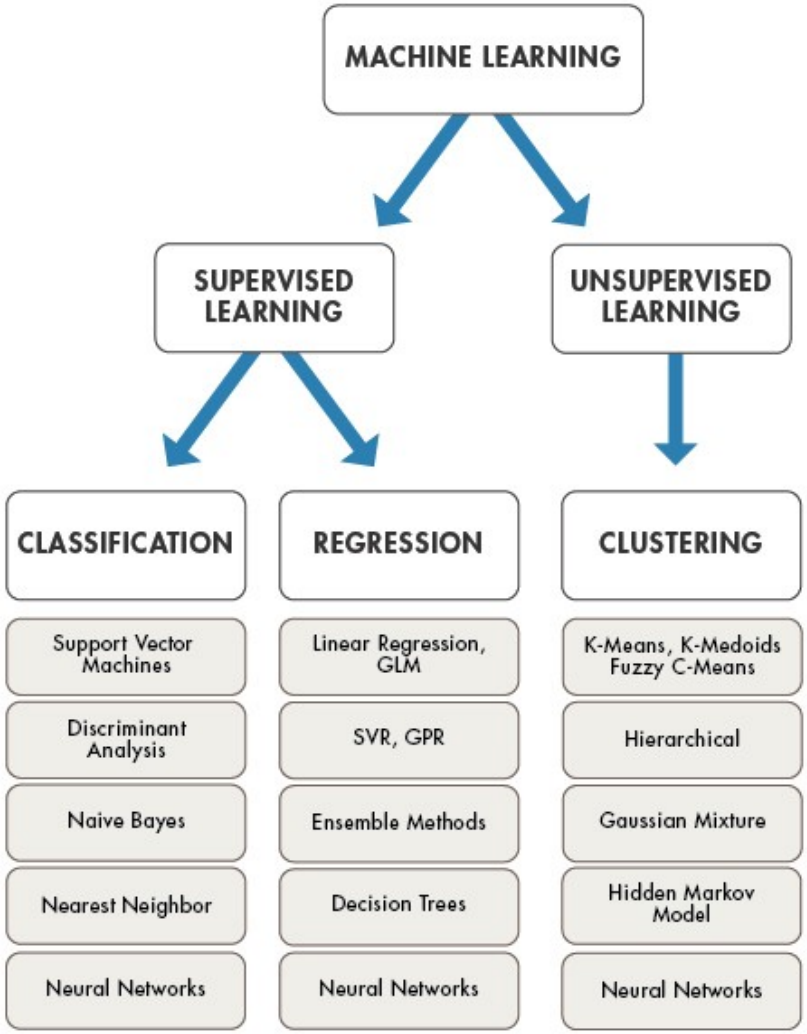
Supervised Machine Learning



Unsupervised Machine Learning



Typical Algorithms in Machine Learning



Application: Supervised

- Predictive analysis based on regression or categorical classification
- Spam detection
- Pattern detection
- Natural Language Processing
- Sentiment analysis
- Automatic image classification
- Automatic sequence processing (for example, music or speech)

Application: Unsupervised

- Object segmentation (for example, users, products, movies, songs, and
- so on)
- Similarity detection
- Automatic labeling

Reinforcement Learning

- Reinforcement Learning is defined as a Machine Learning method that is concerned with how software agents should take actions in an environment.
- Reinforcement Learning is a part of the deep learning method that helps you to maximize some portion of the cumulative reward.

Reinforcement Learning

- Imagine someone playing a video game. The player is the agent, and the game is the environment. The rewards the player gets (i.e. beat an enemy, complete a level), or doesn't get (i.e. step into a trap, lose a fight) will teach him how to be a better player.
- In supervised learning, for example, each decision taken by the model is independent, and doesn't affect what we see in the future.
- In reinforcement learning, instead, we are interested in a long term strategy for our agent, which might include sub-optimal decisions at intermediate steps, and a trade-off between exploration (of unknown paths), and exploitation of what we already know about the environment.

Machine Learning Matters

- What does learning exactly mean? Simply, we can say that learning is the ability to change according to external stimuli and remembering most of all previous experiences.
- So machine learning is an engineering approach that gives maximum importance to every technique that increases or improves the propensity for changing adaptively.

Machine Learning Matters

- A mechanical watch, for example, is an extraordinary artifact, but its structure obeys stationary laws and becomes useless if something external is changed.
- This ability is peculiar to animal and, in particular, to human beings; according to Darwin’s theory, it’s also a key success factor for the survival and evolution of all species. Machines, even if they don't evolve autonomously, seem to obey the same law.

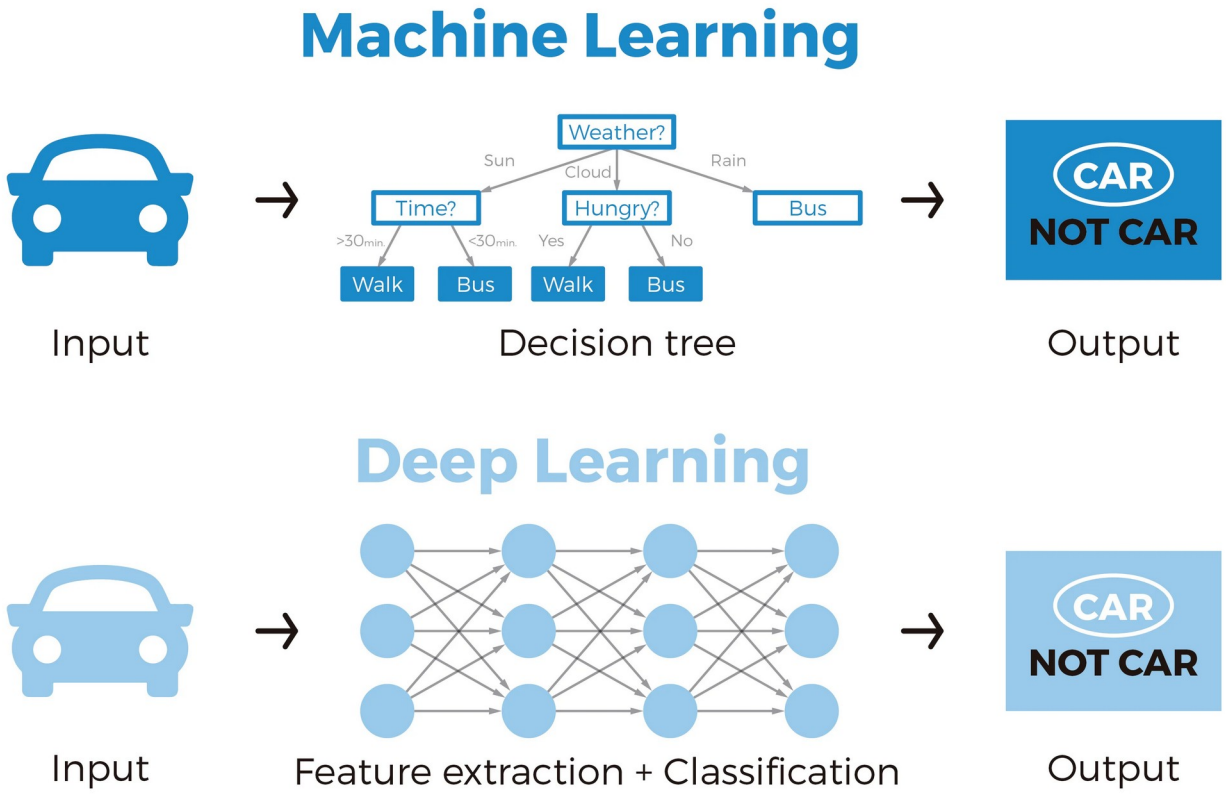
Deep Learning

- Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning.
- Learning can be supervised, semi-supervised or unsupervised.

Deep Learning

- Deep-learning architectures such as deep neural networks, deep belief networks, graph neural networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance

Deep Learning Applications



Deep Learning Applications

- Image classification
- Real-time visual tracking
- Autonomous car driving
- Logistic optimization
- Bioinformatics
- Speech recognition

Bio-inspired computing

- Bio-inspired computing, short for biologically inspired computing, is a field of study which seeks to solve computer science problems using models of biology.
- It relates to connectionism, social behavior, and emergence. Within computer science, bio-inspired computing relates to artificial intelligence and machine learning.
- Bio-inspired computing is a major subset of natural computation.

Bio-inspired computing

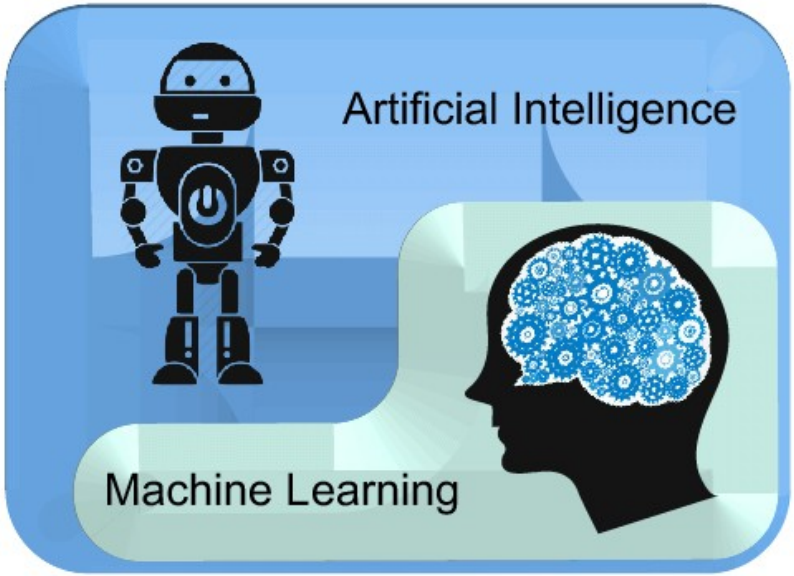
Bio-Inspired Computing Topic	Biological Inspiration
Genetic Algorithms	Evolution
Biodegradability prediction	Biodegradation
Cellular Automata	Life
Emergence	Ants, termites, bees, wasps
Neural networks	The brain
Artificial life	Life
Artificial immune system	Immune system
Rendering (computer graphics)	Patterning and rendering of animal skins, bird feathers, mollusk shells and bacterial colonies
Lindenmayer systems	Plant structures
Communication networks and communication protocols	Epidemiology
Membrane computers	Intra-membrane molecular processes in the living cell
Excitable media	Forest fires, "the wave", heart conditions, axons
Sensor networks	Sensory organs
Learning classifier systems	Cognition, evolution

Learning vs. Designing

- Artificial intelligence and machine learning are the part of computer science that are correlated with each other.
- These two technologies are the most trending technologies which are used for creating intelligent systems.
- Although these are two related technologies and sometimes people use them as a synonym for each other, but still both are the two different terms in various cases.

Learning vs. Designing

- AI is a bigger concept to design intelligent machines that can simulate human thinking capability and behavior, whereas, machine learning is an application or subset of AI that allows machines to learn from data without being programmed explicitly.



Artificial Intelligence

- Artificial intelligence is a field of computer science which makes a computer system that can mimic human intelligence.
- It is comprised of two words "Artificial" and "intelligence", which means "a human-made thinking power."
- Hence we can define it as,
  - Artificial intelligence is a technology using which we can create intelligent systems that can simulate human intelligence.

Artificial Intelligence

- The Artificial intelligence system does not require to be pre-programmed, instead of that, they use such algorithms which can work with their own intelligence.
- It involves machine learning algorithms such as Reinforcement learning algorithm and deep learning neural networks.
- AI is being used in multiple places such as Siri, Google?s AlphaGo, AI in Chess playing, etc.
- Based on capabilities, AI can be classified into three types:
  - Weak AI
  - General AI
  - Strong AI

Artificial Intelligence

- Bio-Inspired computing can be distinguished from traditional artificial intelligence by its approach to computer learning.
- Bio-inspired computing uses an evolutionary approach, while traditional A.I. uses a 'creationist' approach. Bio-inspired computing begins with a set of simple rules and simple organisms which adhere to those rules.
- Over time, these organisms evolve within simple constraints. This method could be considered bottom-up or decentralized.
- In traditional artificial intelligence, intelligence is often programmed from above: the programmer is the creator, and makes something and imbues it with its intelligence.

Brain Inspired Computing

- Brain-inspired computing refers to computational models and methods that are mainly based on the mechanism of the brain, rather than completely imitating the brain.
- The goal is to enable the machine to realize various cognitive abilities and coordination mechanisms of human beings in a brain-inspired manner, and finally achieve or exceed Human intelligence level.

Data All Around

- Lots of data is being collected and war
  - Web data, e-commerce
  - Financial transactions, bank/credit transactions
  - Online trading and purchasing
  - Social Network
  - Cloud

Data and Big Data

- “90% of the world’s data was generated in the last few years.”
- Due to the advent of new technologies, devices, and communication means like social networking sites, the amount of data produced by mankind is growing rapidly every year.
- The amount of data produced by us from the beginning of time till 2003 was 5 billion gigabytes. If you pile up the data in the form of disks it may fill an entire football field.
- The same amount was created in every two days in 2011, and in every six minutes in 2016. This rate is still growing enormously.

Big Data Definition

- No single standard definition…

"Big Data" is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it…

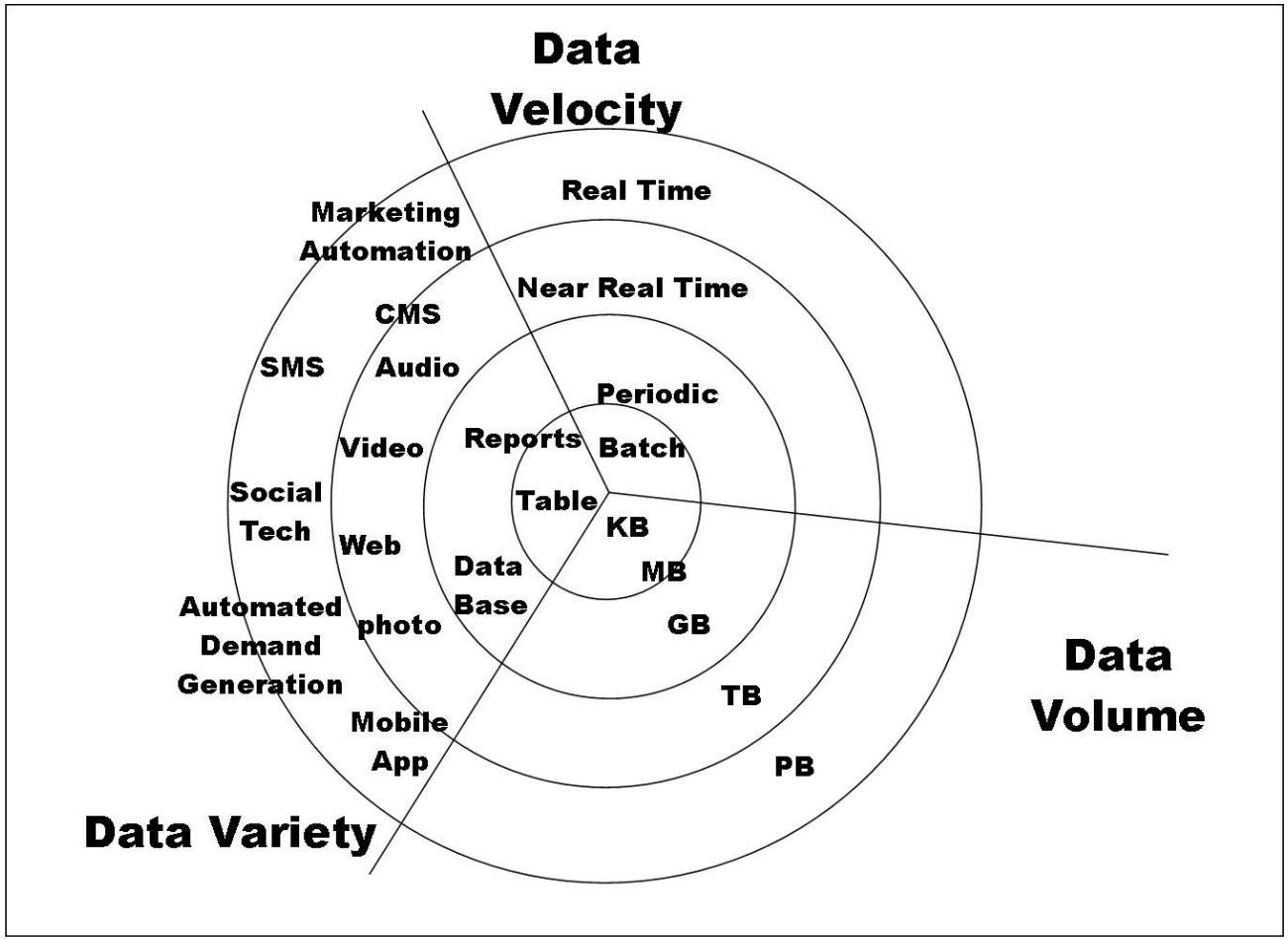
What is Big Data

- Big Data is a collection of large datasets that cannot be processed using traditional computing techniques.
- It is not a single technique or a tool, rather it involves many areas of business and technology.

Big Data

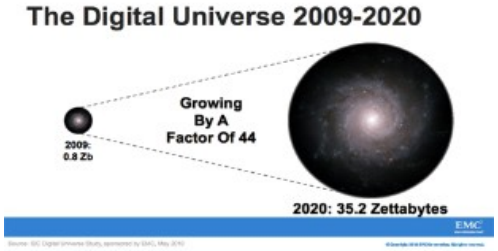
- Big Data is any data that is expensive to manage and hard to extract value from
  - Volume
    - The size of the data
  - Velocity
    - The latency of data processing relative to the growing demand for interactivity
  - Variety and Complexity
    - The diversity of sources, formats, quality, structures.

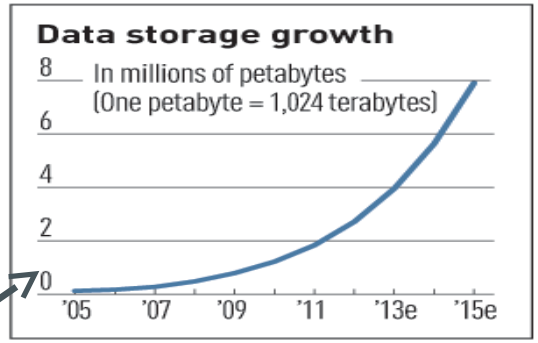
Big Data

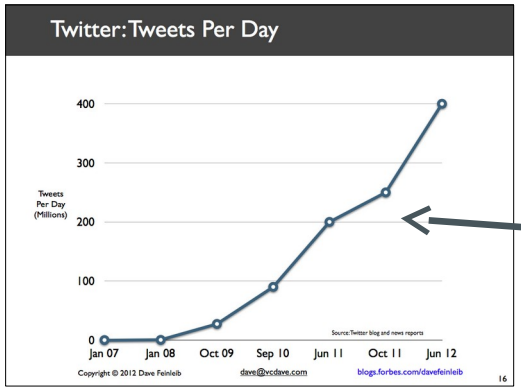

Data
Velocity
Real Time
Near Real Time
Periodic
Batch
Marketing
Automation
CMS
SMS
Audio
Video
Reports
Table
Social
Tech
Web
KB
MB
GB
TB
PB
Data
Base
Automated
Demand
Generation
photo
Mobile
App
Data Variety
Data
Volume

Characteristics of Big Data: Volume

- **Data Volume**
  - 44x increase from 2009 2020
  - From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially





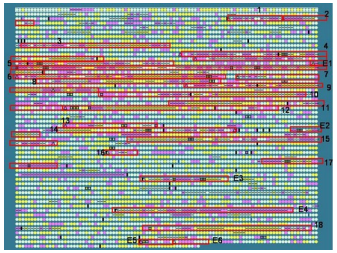


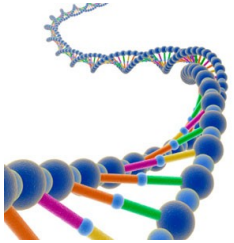
Exponential increase in collected/generated data

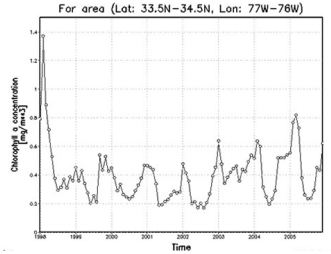
Computer Memory Units

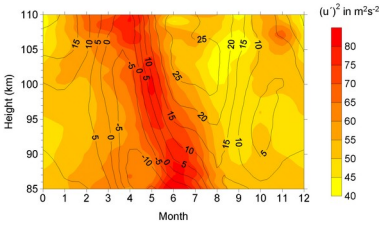
Characteristics of Big Data: Variety

- Various formats, types, and structures
- Text, numerical, images, audio, video, sequences, time series, social media data, multi-dim arrays, etc...
- Static data vs. streaming data
- A single application can be generating/collecting many types of data









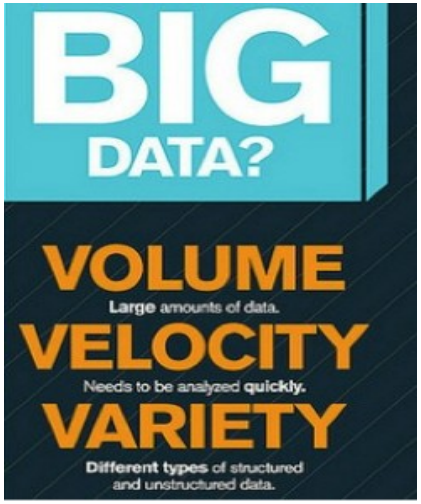
To extract knowledge all these types of data need to linked together

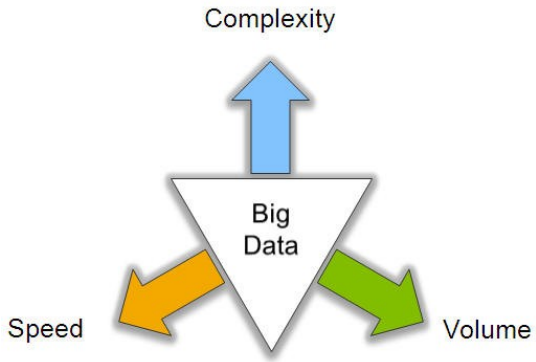
Characteristics of Big Data: Velocity

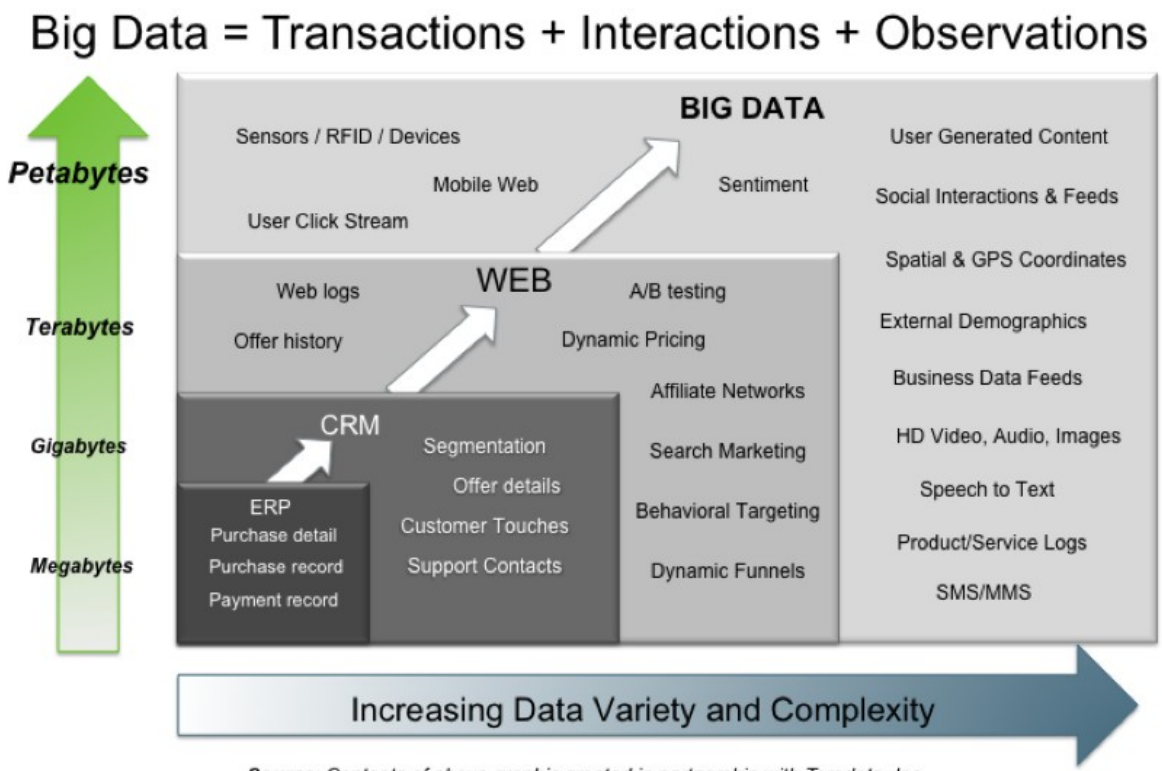
- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions, missing opportunities

- **Examples**
  - **E-Promotions:** Based on your current location, your purchase history, what you like send promotions right now for store next to you.
  - **Healthcare monitoring:** sensors monitoring your activities and body any abnormal measurements require immediate reaction.

Big Data: 3 Vs

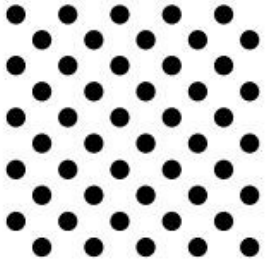
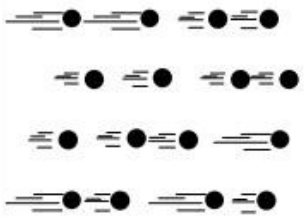
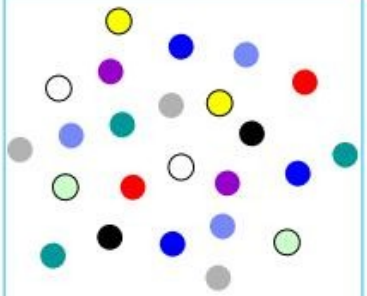
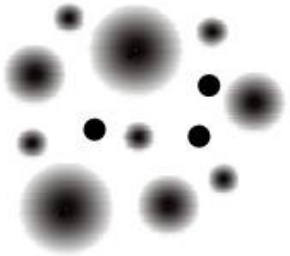






***Source**: Contents of above graphic created in partnership with Teradata, Inc.*

Big Data: The 4th V

Volume	Velocity	Variety	Veracity*
			
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

What Comes Under Big Data?

- Black Box Data: It is a component of helicopter, airplanes, and jets, etc. It captures voices of the flight crew, recordings of microphones and earphones, and the performance information of the aircraft.
- Social Media Data: Social media such as Facebook and Twitter hold information and the views posted by millions of people across the globe.
- Stock Exchange Data: The stock exchange data holds information about the ‘buy’ and ‘sell’ decisions made on a share of different companies made by the customers.
- Power Grid Data: The power grid data holds information consumed by a particular node with respect to a base station.

What Comes Under Big Data?

- Transport Data: Transport data includes model, capacity, distance and availability of a vehicle.
- Search Engine Data: Search engines retrieve lots of data from different databases.
- Structured data: Relational data.
- Semi Structured data: XML data.
- Unstructured data: Word, PDF, Text, Media Logs.

Benefits of Big Data

- Using the information kept in the social network like Facebook, the marketing agencies are learning about the response for their campaigns, promotions, and other advertising mediums.
- Using the information in the social media like preferences and product perception of their consumers, product companies and retail organizations are planning their production.
- Using the data regarding the previous medical history of patients, hospitals are providing better and quick service.

Big Data Technologies

- Operational Big data
- Analytical Big data

Operational Big Data

- These include systems like MongoDB that provide operational capabilities for real-time, interactive workloads where data is primarily captured and stored.
- NoSQL Big Data systems are designed to take advantage of new cloud computing architectures that have emerged over the past decade to allow massive computations to be run inexpensively and efficiently.

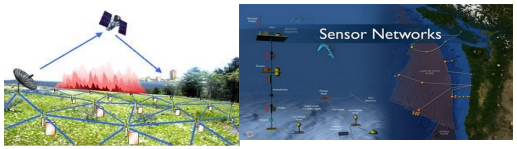
Analytical Big Data

- These includes systems like Massively Parallel Processing (MPP) database systems and MapReduce that provide analytical capabilities for retrospective and complex analysis that may touch most or all of the data.
- MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL, and a system based on MapReduce that can be scaled up from single servers to thousands of high and low end machines.

Who generates Big Data?

Social media and networks
(all of us are generating data)

Scientific instruments
(collecting all sorts of data)



Sensor technology and networks
(measuring all kinds of data)

Mobile devices
(tracking all objects all the time)

Big Data generation models

- **The Model of Generating/Consuming Data has Changed**

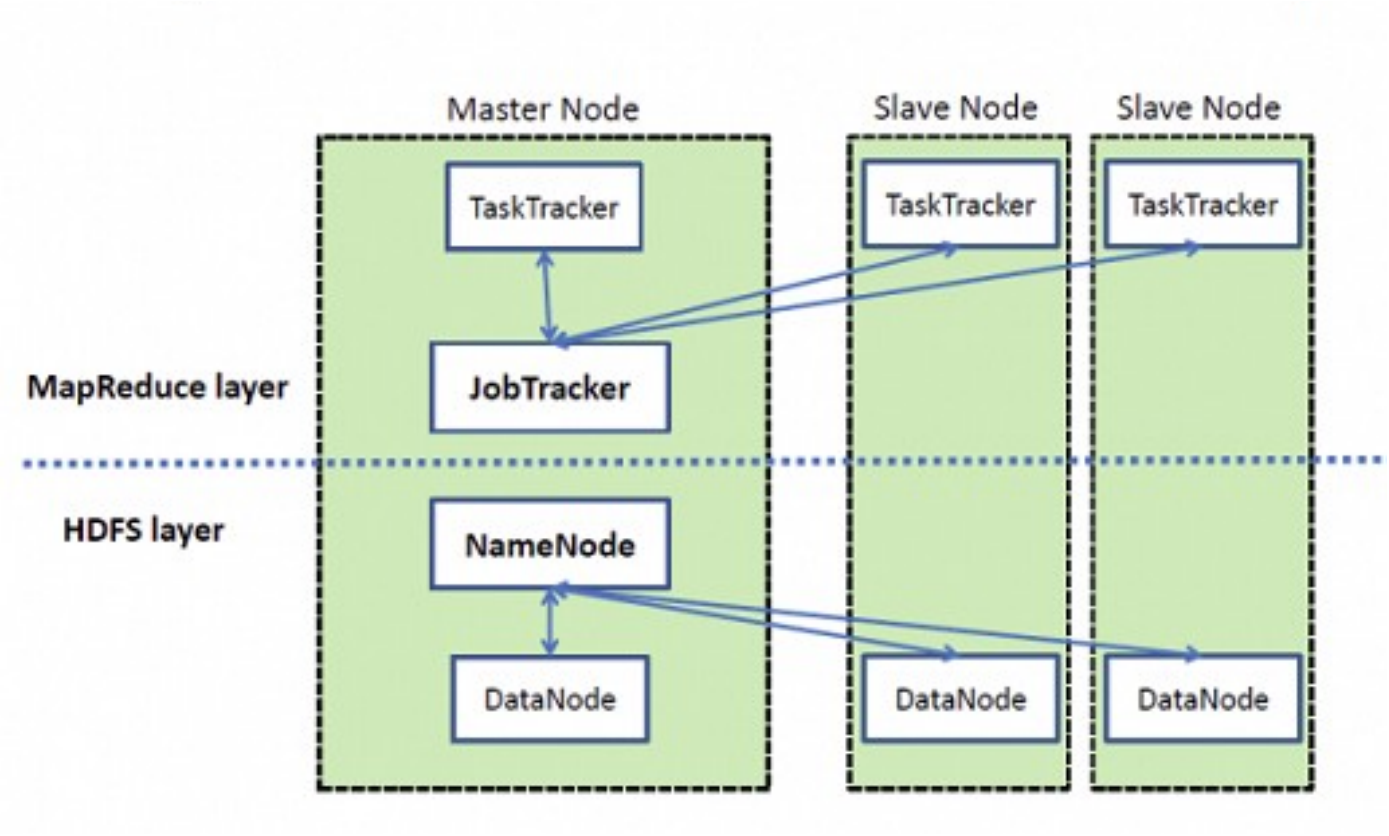
Old Model: Few companies are generating data, all others are consuming data

New Model: all of us are generating data, and all of us are consuming data

Challenges in Big Data

- The major challenges associated with big data are as follows:
  - Capturing data
  - Curation
  - Storage
  - Searching
  - Sharing
  - Transfer
  - Analysis
  - Presentation

Hadoop


Master Node
Slave Node
Slave Node
TaskTracker
TaskTracker
TaskTracker
JobTracker
MapReduce layer
HDFS layer
NameNode
DataNode
DataNode
DataNode

Data: Types of variables

- Dependent and Independent variables
- Continuous and Categorical variables

Dependent and independent

- An independent variable, sometimes called an **experimental** or **predictor** variable, is a variable that is being manipulated in an experiment in order to observe the effect on a dependent variable, sometimes called an **outcome** or **response** variable.
- Imagine that a tutor asks 100 students to complete a maths test. The tutor wants to know why some students perform better than others. Whilst the tutor does not know the answer to this, she thinks that it might be because of two reasons: (1) some students spend more time revising for their test; and (2) some students are naturally more intelligent than others. As such, the tutor decides to investigate the effect of revision time and intelligence on the test performance of the 100 students.
  - Dependent Variable: Test Mark (measured from 0 to 100)
  - Independent Variables: Revision time (measured in hours) Intelligence (measured using IQ score)

Categorical and Continuous

- Categorical variables are also known as discrete or qualitative variables.
  - Categorical variables can be further categorized as either nominal, ordinal or dichotomous.
- Continuous variables are also known as quantitative variables.
  - Continuous variables can be further categorized as either interval or ratio variables.

Categorical variables

- Nominal variables
  - are variables that have two or more categories, but which do not have an intrinsic order.
- Dichotomous variables
  - are nominal variables which have only two categories or levels.
- Ordinal variables
  - are variables that have two or more categories just like nominal variables only the categories can also be ordered or ranked.

Continuous variables

- Interval variables
  - are variables for which their central characteristic is that they can be measured along a continuum and they have a numerical value (for example, temperature measured in degrees Celsius or Fahrenheit).
- Ratio variables
  - are interval variables, but with the added condition that 0 (zero) of the measurement indicates that there is none of that variable.

Data formats

- In a supervised learning problem, there will always be a dataset, defined as a finite set of real vectors with m features each:

$$X = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\} \, where \, \bar{x}_i \in \mathbb{R}^m$$

- Considering that our approach is always probabilistic, we need to consider each X as drawn from a statistical multivariate distribution D.
- For our purposes, it's also useful to add a very important condition upon the whole dataset X: we expect all samples to be independent and identically distributed (i.i.d).

Data Formats

- The corresponding output values can be both numerical-continuous or categorical. In the first case, the process is called regression, while in the second, it is called classification. Examples of numerical outputs are:

$$Y = \{y_1, y_2, \ldots, y_n\} \text{ where } y_n \in (0,1) \text{ or } y_i \in \mathbb{R}^+$$

- Categorical examples are:

$$y_i \in \{red, black, white, green\} \text{ or } y_i \in \{0,1\}$$

Data Formats

- We define generic regressor, a vector-valued function which associates an input value to a continuous output and generic classifier, a vector-values function whose predicted output is categorical (discrete).
- If they also depend on an internal parameter vector which determines the actual instance of a generic predictor, the approach is called parametric learning.

Learnability

- A parametric model can be split into two parts: a static structure and a dynamic set of parameters.
- The former is determined by choice of a specific algorithm and is normally immutable (except in the cases when the model provides some re-modeling functionalities), while the latter is the objective of our optimization.

Learnability

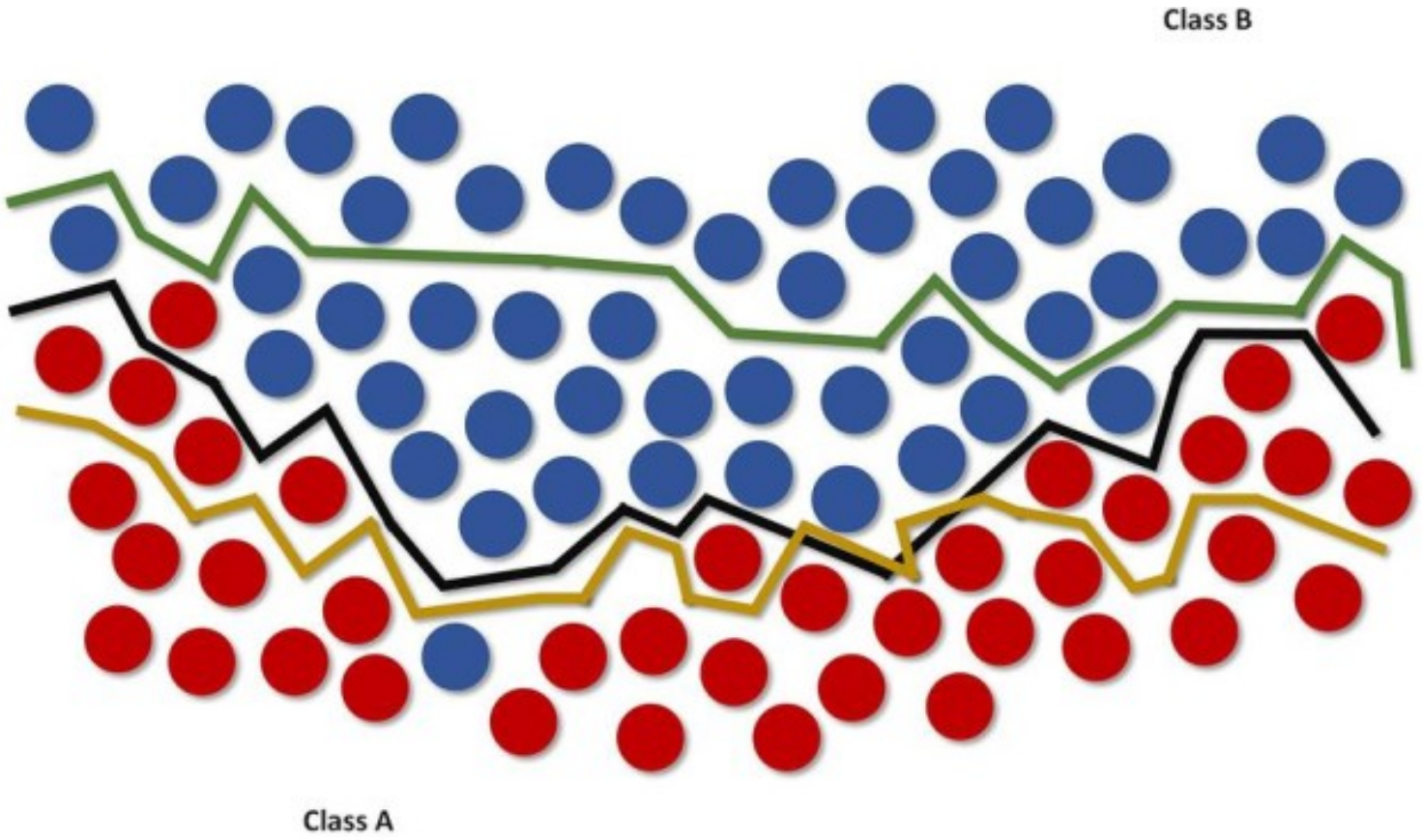
- Considering n unbounded parameters, they generate an n-dimensional space (imposing bounds results in a sub-space without relevant changes in our discussion) where each point, together with the immutable part of the estimator function, represents a learning hypothesis H (associated with a specific set of parameters):

$$H = \{\theta_1, \theta_2, \ldots, \theta_n\}$$

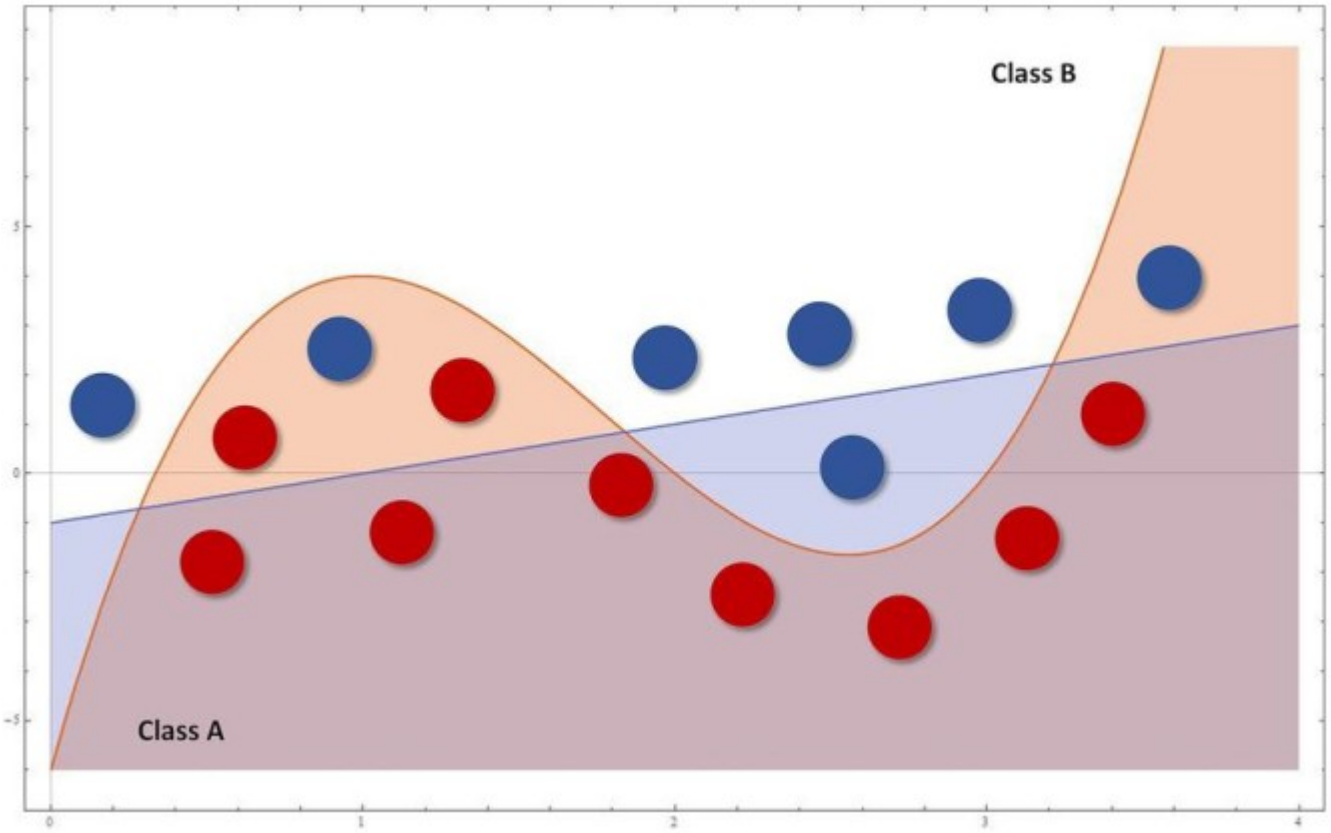
Learnability

- The goal of a parametric learning process is to find the best hypothesis whose corresponding prediction error is minimum and the residual generalization ability is enough to avoid overfitting.
- In the following figure, there's an example of a dataset whose points must be classified as red (Class A) or blue (Class B).
- Three hypotheses are shown: the first one (the middle line starting from left) misclassifies one sample, while the lower and upper ones misclassify 13 and 23 samples respectively:

Learnability



Learnability: Another Scenario



Statistical learning approaches

- Imagine that you need to design a spam-filtering algorithm starting from this initial (over-simplistic) classification based on two parameters:

Parameter	**Spam emails (X_1)**	**Regular emails ($X2$)**
p_1 - Contains > 5 blacklisted words	80	20
p_2 - Message length < 20 characters	75	25

Statistical learning approaches

- We have collected 200 email messages (X) (for simplicity, we consider p_1 and p_2 mutually exclusive) and we need to find a couple of probabilistic hypotheses (expressed in terms of p 1 and p 2), to determine:

$$P(spam|h_{p1}, h_{p2})$$

- We also assume the conditional independence of both terms (it means that h p1 and h p2 contribute conjunctly to spam in the same way as they were alone).

Bayes Theorem

- Bayes Theorem: Principled way of calculating a conditional probability without the joint probability. It is often the case that we do not have access to the denominator directly, e.g. P(B).
- We can calculate it an alternative way; for example:
  - P(B) = P(B|A) * P(A) + P(B|not A) * P(not A)
- This gives a formulation of Bayes Theorem that we can use that uses the alternate calculation of P(B), described below:
  - P(A|B) = P(B|A) * P(A) / P(B|A) * P(A) + P(B|not A) * P(not A)

Bayes Theorem

- Firstly, in general, the result P(A|B) is referred to as the posterior probability and P(A) is referred to as the prior probability.
  - P(A|B): Posterior probability.
  - P(A): Prior probability.
- Sometimes P(B|A) is referred to as the likelihood and P(B) is referred to as the evidence.
  - P(B|A): Likelihood.
  - P(B): Evidence.
- This allows Bayes Theorem to be restated as:
  - Posterior = Likelihood * Prior / Evidence

Elements of information theory

- A machine learning problem can also be analyzed in terms of information transfer or exchange.
- Our dataset is composed of n features, which are considered independent (for simplicity, even if it's often a realistic assumption) drawn from n different statistical distributions.
- Therefore, there are n probability density functions $p_i(x)$ which must be approximated through other n $q_i(x)$ functions.

Elements of information theory

- The most useful measure is called entropy:

$$H(X) = -\sum_{x \in X} p(x) log_2 p(x)$$

- This value is proportional to the uncertainty of X and it's measured in bits (if the logarithm has another base, this unit can change too).
- For many purposes, a high entropy is preferable, because it means that a certain feature contains more information.

Elements of information theory

- For example, in tossing a coin (two possible outcomes), *H(X) = 1* bit, but if the number of outcomes grows, even with the same probability, *H(X)* also does because of a higher number of different values and therefore increased variability.
- It's possible to prove that for a Gaussian distribution (using natural logarithm):

$$H(X) = \frac{1}{2}(1 + \ln(2\pi\sigma^2))$$

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com